

Обработка языка. Практика 2

Векторное представление слов

Михаил Пожидаев

19 сентября 2023 г.

Вычисление

Вывод вариантов с расстоянием

```
import gensim
pos = ["гвоздь_NOUN"]
neg = ["шляпка_NOUN"]
word2vec = gensim.models.KeyedVectors
    .load_word2vec_format('cbow.txt', binary=False)
dist = word2vec.most_similar(positive=pos, negative=neg)
for i in dist:
    print(i)
```

Фильтрация

Вывод только имён существительных

```
import re
pat = re.compile("(*)_NOUN")
...
for i in dist:
    e = pat.match(i[0])
    if e is not None:
        print(e.group(1))
```

Спасибо за внимание!

Всё о курсе: <https://marigostra.ru/materials/nlp.html>

E-mail: msp@luwrain.org

Канал в Телеграм: <https://t.me/MarigostraRu>